Medical Image Analysis 69 (2021) 101951

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Automated ultrasound assessment of amniotic fluid index using deep learning

Hyun Cheol Cho[a], Siyu Sun[a], Chang Min Hyun[a], Ja-Young Kwon[b], Bukweon Kim[a], Yejin Park[b,*], Jin Keun Seo[a]

[a] School of Mathematics and Computing (Computational Science and Engineering), Yonsei University, Seoul 03722, Republic of Korea
[b] Department of Obstetrics and Gynecology, Institute of Womens Life Medical Science, Yonsei University College of Medicine, Seoul 03722, Republic of Korea

ARTICLE INFO

Article history:
Received 27 February 2020
Revised 13 December 2020
Accepted 21 December 2020
Available online 7 January 2021

Keywords:
Amniotic fluid index
Ultrasound image
Image segmentation
Deep learning

ABSTRACT

The estimation of antenatal amniotic fluid (AF) volume (AFV) is important as it offers crucial information about fetal development, fetal well-being, and perinatal prognosis. However, AFV measurement is cumbersome and patient specific. Moreover, it is heavily sonographer-dependent, with measurement accuracy varying greatly depending on the sonographer's experience. Therefore, the development of accurate, robust, and adoptable methods to evaluate AFV is highly desirable. In this regard, automation is expected to reduce user-based variability and workload of sonographers. However, automating AFV measurement is very challenging, because accurate detection of AF pockets is difficult owing to various confusing factors, such as reverberation artifact, AF mimicking region and floating matter. Furthermore, AF pocket exhibits an unspecified variety of shapes and sizes, and ultrasound images often show missing or incomplete structural boundaries. To overcome the abovementioned difficulties, we develop a hierarchical deep-learning-based method, which consider clinicians' anatomical-knowledge-based approaches. The key step is the segmentation of the AF pocket using our proposed deep learning network, AF-net. AF-net is a variation of U-net combined with three complementary concepts - atrous convolution, multi-scale side-input layer, and side-output layer. The experimental results demonstrate that the proposed method provides a measurement of the amniotic fluid index (AFI) that is as robust and precise as the results from clinicians. The proposed method achieved a Dice similarity of 0.877 ± 0.086 for AF segmentation and achieved a mean absolute error of 2.666 ± 2.986 and mean relative error of 0.018 ± 0.023 for AFI value. To the best of our knowledge, our method, for the first time, provides an automated measurement of AFI.

1. Introduction

Amniotic fluid (AF) is essential to fetal development, as AF is required for promotion of maturation of the lungs, gastrointestinal development, and musculoskeletal system development (Manning et al., 1980). Amniotic fluid volume (AFV) is deemed abnormal when it lies outside the range of 500–2000 ml. Oligohydramnios is defined as less than 500 ml of the total volume whereas hydramnios is defined when the total volume is greater than 2000 ml (Beall et al., 2007; Moore and Cayle, 1990). Oligohydramnios contributes to perinatal morbidity, including increased risk of cesarean delivery for fetal distress, low Apgar score, neonatal intensive care unit admission, low-birth-weight neonate, and meconium aspiration syndrome (Rutherford et al., 1987; Rabie et al., 2017). Furthermore, oligohydramnios from early gestation could result in pulmonary hypoplasia (Wu et al., 2017a). There is an increased risk of preterm birth, perinatal mortality, low Apgar score, admission to neonatal intensive care unit, fetal distress, and neurodevelopmental abnormalities in pregnancies with hydramnios (Magann et al., 2007). Reported causes of hydramnios include fetal structural abnormalities, genetic disorders, fetal anemia, and placental tumors, multiple gestation and maternal diabetes, most of which are idiopathic (Pri-Paz et al., 2012; Yefet and Daniel-Spiegel, 2016).

AFV is an important indicator that reflects fetal well-being as well as perinatal prognosis (Dubil and Magann, 2013; Manning et al., 1980; of Obstetricians and Gynecologists, 2014). Estimation of AFV is an important part of fetal surveillance as a component of the biophysical profile. In particular, it is a critical factor for deter-

* Corresponding author.
E-mail address: DRYJPARK02@yuhs.ac.kr (Y. Park).

https://doi.org/10.1016/j.media.2020.101951

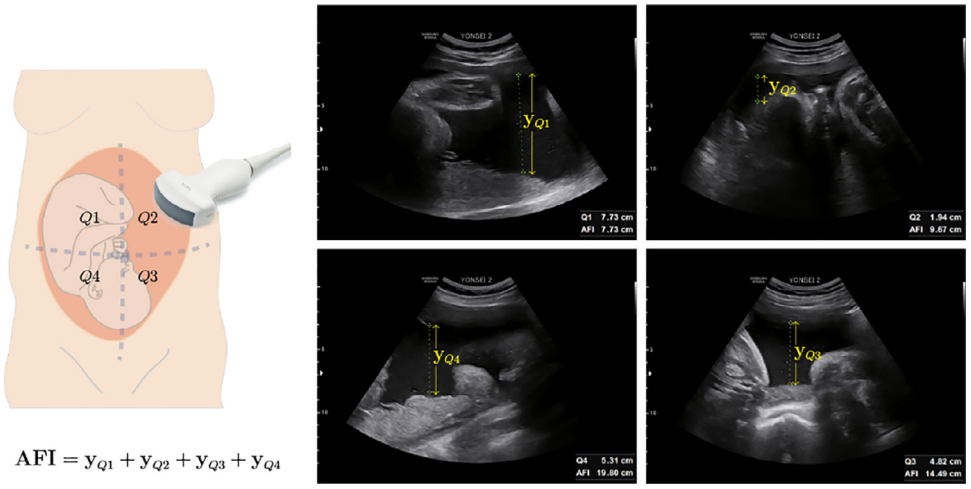


Fig. 1. AFI measurement. AFI is the sum of deepest depth of amniotic fluid in each quadrant of ultrasound images. Here, Q1, Q2, Q3, and Q4 indicate each quadrant and $\mathbf{y}_{Q1}$, $\mathbf{y}_{Q2}$, $\mathbf{y}_{Q3}$, and $\mathbf{y}_{Q4}$ indicate the deepest depth of amniotic fluid in each quadrant, respectively.

mining prompt deliveries (Chamberlain et al., 1984). The fact that hydramnios or oligohydramnios can affect the accuracy of biometry also adds to the clinical importance of accurate AF measurement (Rutherford et al., 1987; Magann et al., 2011). Therefore, antenatal AFV estimation is a fundamental measurement required during antenatal ultrasound (of Obstetricians et al., 2016).

Ultrasound (US) is widely used to estimate AFV non-invasively, and AFV is usually measured by evaluating the four-quadrant amniotic fluid index (AFI) or the single deep vertical pocket (SDP) technique (Phelan et al., 1987a; 1987b; Kehl et al., 2016; Luntsi et al., 2019), which can be seen in Fig. 1. Oligohydramnios is considered when less than 5 cm of AFI (or less than 2 cm of SDP) is observed, while hydramnios is considered when more than 24 cm of AFI (or more than 8 cm of SDP) is observed. To measure the AFI or SDP, sonographers manually follow the following time-consuming steps, which involves multiple keystrokes and probe motions: *(i) The clinicians determine a proper AF pocket and then (ii) measure the depth of the AF by estimating the suitable point.* Although AFI and SDP are known to be reproducible and semi-quantitative, the manual AFI measurement is highly dependent on the sonographer's experience in practice.

This paper aims to solve these problems by providing an automatic method to measure AFI that is operator independent while maintaining a certain accuracy level. To the best of our knowledge, our study appears to be the first attempt of this challenging task, which is unlike other ultrasound-based automatic fetal biometry measurements (Jang et al., 2017; Wu et al., 2017b; Kim et al., 2018; 2019; Liu et al., 2019). The difficulty arises from various confusing factors, such as reverberation artifact, AF mimicking region, and floating matter. Moreover, AF pockets show an unspecified variety of shapes and sizes, and ultrasound images often include missing or incomplete structural boundaries, as shown in Fig. 2.

To overcome these difficulties, we developed a deep-learning-based method that uses an anatomical knowledge-based hierarchical process to reflect the procedure of a clinician's measurement of AFI. The framework of our method can be divided into two steps. First, we segmented the AF pocket with our proposed network called AF-net to find a proper AF pocket. Second, the segmented result is used to find deepest vertical line of AF pocket.

Our proposed AF-net is essentially U-net (Ronneberger et al., 2015) with three additional concepts to overcome the above-mentioned difficulties; (i) atrous convolution (Chen et al., 2017a), (ii) multi-scale side-input layer, (iii) side-output (Mehta and Sivaswamy, 2017). U-net is a state-of-the art medical image segmentation technique, providing useful results in various areas even with relatively small amount of data (Kim et al., 2019; Yap et al., 2017; Dong et al., 2017; Hyun et al., 2019; 2018). However, U-net has shown several limitations during AF segmentation, particularly when the ultrasound image is corrupted by various confusing factors or when there are large differences in size among the AF pockets. To overcome these limitations, we integrated the three mentioned complementary techniques with U-net.

The efficacy of the proposed method was experimentally validated using 125 clinical US images. We observed that the AFI measurements made by the proposed method is significantly comparable to those obtained manually by clinicians. Our proposed method successfully measures the AFI even for ultrasound images with various confusing factors. Quantitatively, the method achieved a Dice metric of 0.877 ± 0.086 for amniotic fluid segmentation with a mean absolute error of 2.666 ± 2.986 mm and a mean relative error of 0.018 ± 0.023 for the AFI value.

2. Methods

Our goal is to find a map $f : \boldsymbol{x}_o \mapsto \boldsymbol{y}$, where $\boldsymbol{x}_o$ represents the ultrasound image and $\boldsymbol{y}$ is the corresponding AFI. The map f takes the clinicians' anatomical knowledge-based approach into account, reflecting the procedure of their measurement of AFI. The proposed framework, described in Fig. 3, is divided into the following two steps: (i) segmentation of the AF pocket with a preprocessed image $\boldsymbol{x}$; and (ii) measurement of AFI using the result obtained from the first step.

For a robust and accurate segmentation of AF pocket, we developed a deep learning method for finding a semantic segmentation function $f_{seg} : \boldsymbol{x} \mapsto \mathbf{T}$ that maps from a preprocessed image $\boldsymbol{x}$ to binary images $\mathbf{T} = (T^{(1)}, T^{(2)})$, where $T^{(1)}$ and $T^{(2)}$ denote segmented binary images of the AF pocket and the remaining region, respectively. The deep learning-based method uses labeled training data $\{(\boldsymbol{x}_j, \mathbf{T}_j) : j = 1, \ldots, N\}$ to learn f_{seg}, which can be achieved by

$$f_{seg} = \underset{f_{seg} \in \mathbb{NN}}{\operatorname{argmin}} \frac{1}{N} \sum_{j=1}^{N} \mathbb{L}_{seg}(f_{seg}(\boldsymbol{x}_j), \mathbf{T}_j), \tag{1}$$

where $\mathbb{NN}$ is the set of segmentation functions described by a special form of neural network, $\mathbb{L}_{seg}$ is a loss function for the deep learning neural network, and N is the number of training data.

The proposed network, called AF-net, is motivated from three popular convolution neural network(CNN) structures: U-net

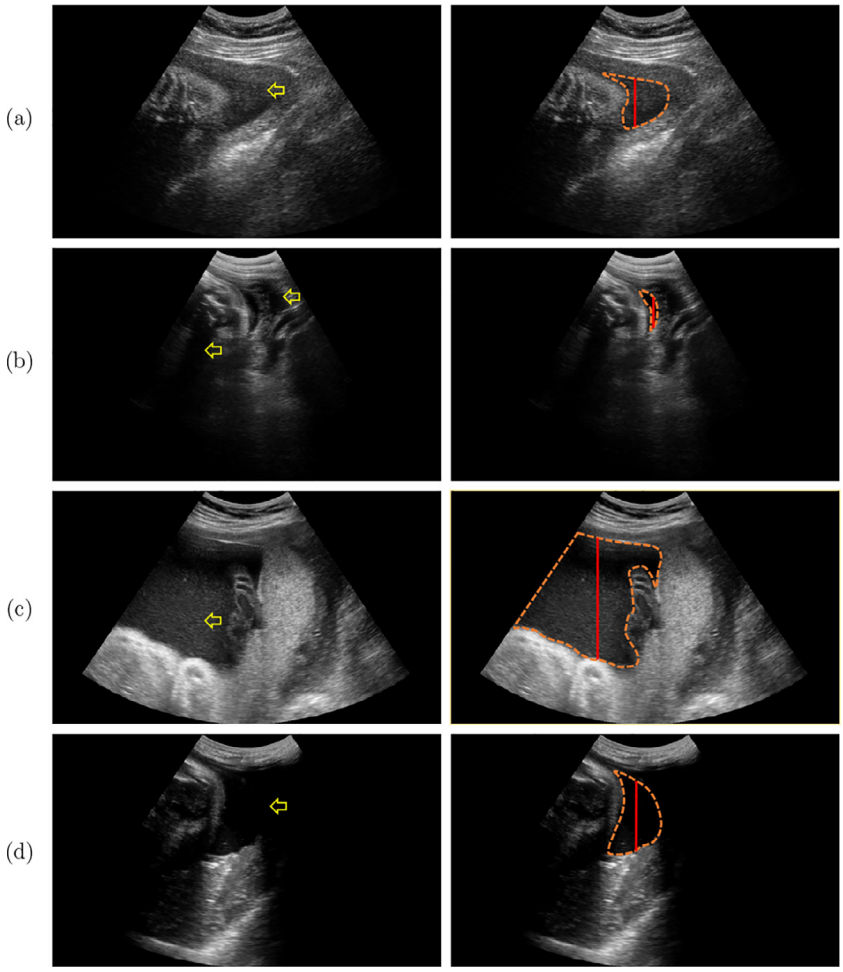


Fig. 2. Confusing factors in segmenting AF pocket: (a) reverberation artifact, (b) AF mimicking region, (c) floating matters, and (d) incomplete or missing boundary. The reason why these obstacles are difficult to deal with is that there is. no specific morphological feature and size in AF pocket. Here, the orange dashed lines indicate the ground true of AF pocket, red solid lines indicate AFI lines and yellow arrows indicate the factors to misrecognition. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

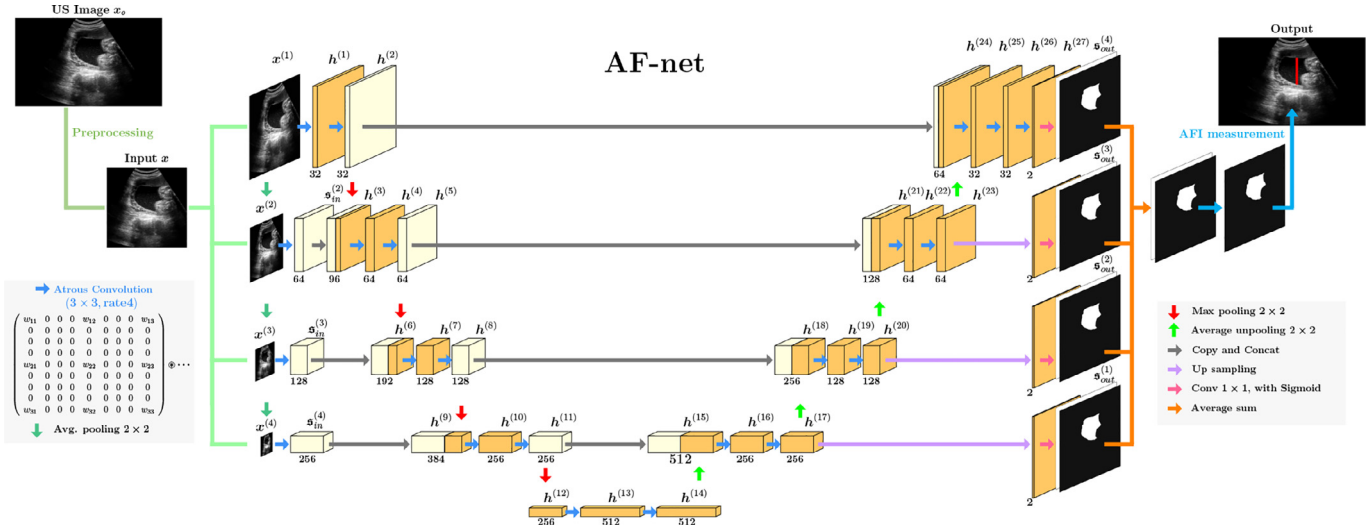


Fig. 3. Overall structure of our proposed method. The deep semantic segmentation network, AF-net, is proposed to segment AF pocket. Our AF-net architecture consists of encoding and decoding path, multi-scale input layer, side-output layer, and atrous convolution. This architecture allows to effectively deal with the challenging factors described in Fig. 2. After segmentation using AF-net, we perform AFI measurement.

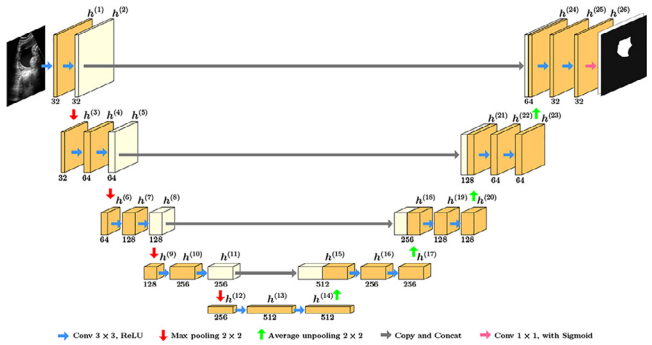


Fig. 4. U-net architecture for semantic segmentation of AF pocket. This U-net approach shows limited performance in dealing with the challenging factors in Fig. 2.

(Ronneberger et al., 2015), M-net (Mehta and Sivaswamy, 2017) and DeepLab (Chen et al., 2017a). For ease of explanation, we begin with a description of U-net and its limitation to our task.

2.1. U-net based approach and its limitation to our task

U-net (Ronneberger et al., 2015) is one of the most popular networks for medical image segmentation. U-net consists of two main parts; encoding and decoding path, as described in Fig. 4. At the first convolution layer in the encoding path, a set of feature maps(denoted by $\boldsymbol{h}^{(1)}$) is computed using a set of filters(denoted by $\boldsymbol{W}^{(1)}$)

$$\boldsymbol{h}^{(1)} = ReLU(\boldsymbol{x} \circledast \boldsymbol{W}^{(1)}) \tag{2}$$

where $\circledast$ stands for standard convolution with stride 1 and $ReLU(x) = \max\{x, 0\}$ is the rectified linear unit. Similarly, at the second layer, the second feature map of $\boldsymbol{h}^{(2)}$ is obtained by $\boldsymbol{h}^{(2)} = ReLU(\boldsymbol{h}^{(1)} \circledast \boldsymbol{W}^{(2)})$. After two convolution layers, max pooling operation is applied to reduce the dimensions of the feature maps. Three consecutive processes are repeated up to the end of the encoding path to extract feature maps. The decoding path is a reverse process of the encoding path, replacing the pooling operator by the average un-pooling to restore the size of the output. The un-pooled output is concatenated with the corresponding feature from the encoding path. At the last layer, we take the pixel-wise softmax activation function after applying a 1×1 convolution. We visualize the overall process of U-net in Fig. 4.

With the U-net architecture, f_{seg} is determined from the set of filters $\{\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)}, \ldots\}$ which can be achieved through the optimization process of Eq. (1). In our experience, U-net often misrecognizes the AF pocket because of the following confusing factors; reverberation artifacts, AF mimicking regions, floating matters, and missing or incomplete boundary of AF pocket. As far as we understand, it seems to be mainly caused by the following three reasons: (i) As shown in Fig. 5, the U-net structure has a too small size of receptive field to extract features in the encoding path. (ii) The single level of receptive field size in encoding path of U-net makes it hard to handle the large differences in AF shape properly. (iii) Because of the skip connections in U-net, the number of feature channels in the decoding path significantly increases; therefore, the decoder path seems to propagate useless information (Hu et al., 2019) such as dark regions that resemble AF.

2.2. Main result: AF-net

In order to address the limitations of U-net, our proposed network makes use of the following complementary concepts:

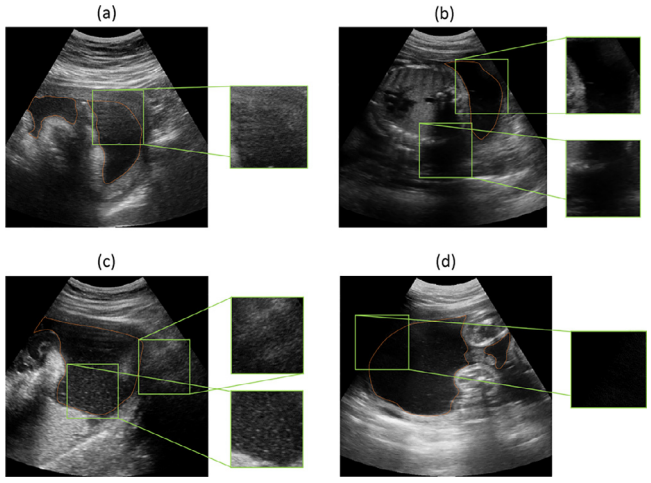


Fig. 5. U-net has relatively too small size of receptive field to deal with the confusing factors; (a) reverberation artifact, (b) AF mimicking region, (c) floating matters, and (d) incomplete or missing boundary. The orange line indicate the ground true of AF pocket, and the green box indicate the maximum receptive field size of U-net in the encoding path. These confusing factors make hard to distinguish a AF pocket by just looking at local region (green box). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

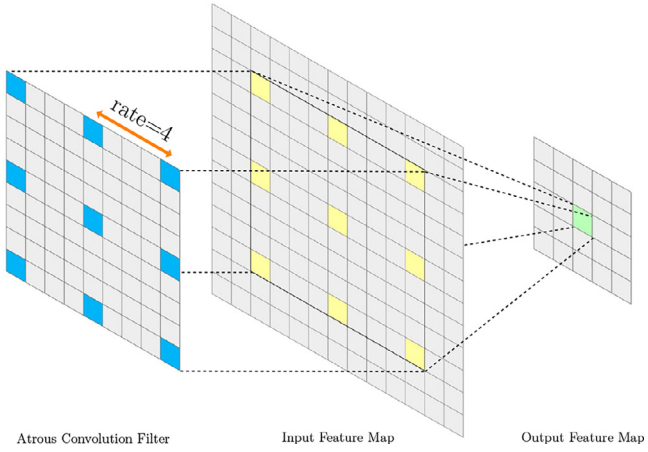


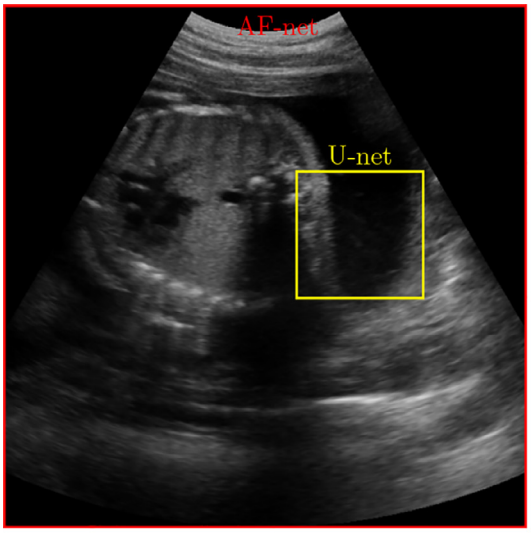
Fig. 6. Atrous convolution of rate 4 with a 3×3 filter. The atrous convolution is equivalent to the standard convolution with the 9×9 kernel, which is upsampled from the 3×3 filter by adding three zeros between two consecutive filter values along each spatial dimension.

2.2.1. Atrous convolution

Atrous convolution (Chen et al., 2017a; 2017b; 2018) with a rate r uses an up-sampled filter produced by inserting $r-1$ zeros between two consecutive filter values along each spatial dimension, as seen in Fig. 6. Atrous convolution of a rate r with $K \times L$ filter is expressed by the standard convolution with $((K-1)r+1) \times ((L-1)r+1)$ filter. For example, the output $\boldsymbol{h}$ of atrous convolution of $\boldsymbol{x}$ with a 3×3 filter $\boldsymbol{W}$ and the factor 2 is expressed by

$$\begin{aligned} \boldsymbol{h} &= \boldsymbol{x} \circledast^2 \underbrace{\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix}}_{\boldsymbol{W}} \\ &= \boldsymbol{x} \circledast \begin{bmatrix} w_{11} & 0 & w_{12} & 0 & w_{13} \\ 0 & 0 & 0 & 0 & 0 \\ w_{21} & 0 & w_{22} & 0 & w_{23} \\ 0 & 0 & 0 & 0 & 0 \\ w_{31} & 0 & w_{32} & 0 & w_{33} \end{bmatrix} \end{aligned} \tag{3}$$

where $\circledast^2$ stands for atrous convolution with rate 2.



(a)

Layer	U-net	AF-net
$h^{(1)}$	3×3	9×9
$h^{(2)}$	5×5	17×17
$h^{(3)}$	6×6	18×18
$h^{(4)}$	10×10	34×34
$h^{(5)}$	14×14	50×50
$h^{(6)}$	16×16	36×36, 52×52
$h^{(7)}$	24×24	68×68, 84×84
$h^{(8)}$	32×32	100×100, 116×116
$h^{(9)}$	36×36	72×72, 104×104, 120×120
$h^{(10)}$	52×52	136×136, 184×184, 168×168
$h^{(11)}$	68×68	200×200, 232×232, 248×248
$h^{(12)}$	76×76	208×208, 240×240, 256×256
$h^{(13)}$	108×108	336×336, 368×368, 384×384
$h^{(14)}$	140×140	464×464, 496×496, 512×512

(b)

Fig. 7. Receptive field size analysis for U-net and AF-net. In (a), we visualize the receptive field associated with one node in the last layer of the encoding path in U-net and AF-net in terms of an input image whose size is 512×512. The entire input US image is covered by the receptive field of AF-net in the last encoding layer, but that of U-net is too small size to recognize the proper AF pocket. The receptive field size associated with a node in each layer of the encoding path is given in (b). AF-net has multiple levels of receptive field sizes in contrast to U-net.

For increasing the receptive field size, adding more pooling operations in the network reduces spatial information and, therefore, may hamper dense prediction of AF pocket, where detailed spatial information is required. Instead using the pooling operators, atrous convolution effectively extracts denser feature maps while keeping certain spatial information (Chen et al., 2018; He et al., 2015). Therefore, atrous convolution is adopted in the encoding path and the decoding path of AF-net, so that the receptive field size becomes considerably larger than that of U-net, as seen in Fig. 7. The enlarged receptive field via atrous convolution is expected to improve segmentation performance by allowing the network to utilize not only the local context but also the global context. Through the enriched global context, our proposed network can segment the suitable AF pocket by overcoming the difficulties arising from various confusing factors.

2.2.2. Multi-scale side-input layer

Inspired by M-net (Mehta and Sivaswamy, 2017), our network adopted the multi-scale side-input layers to achieve multiple level of receptive field sizes. In AF-net, the multi-scale side inputs $\{\boldsymbol{x}^{(i)}\}_{i=2}^{4}$ are achieved by taking average pooling operation to the input image $\boldsymbol{x}$, defined by

$$\boldsymbol{x}^{(i)} = \mathit{Ave\text{-}pool}(\boldsymbol{x}^{(i-1)}, 2), \quad i \in \{2, 3, 4\} \tag{4}$$

where $\mathit{Ave\text{-}pool}(\boldsymbol{x}^{(i)}, 2)$ represents the bilinear down-sampling of the factor 2 to $\boldsymbol{x}^{(i)}$. Here, we denote $\boldsymbol{x}$ by $\boldsymbol{x}^{(1)}$ for notational convenience. Each $\boldsymbol{x}^{(i)}$ generates the corresponding feature map $\mathfrak{s}_{in}^{(i)}$ through atrous convolution of rate 4 with filters whose kernel size is 3×3, given by

$$\mathfrak{s}_{in}^{(i)} = \mathit{ReLU}(\boldsymbol{x}^{(i)} \circledast^{4} \boldsymbol{W}^{(\mathfrak{s}_{in}^{(i)})}), \quad i \in \{2, 3, 4\} \tag{5}$$

Each $\mathfrak{s}_{in}^{(i)}$ is concatenated to the starting point of the encoding path, as observed in Fig. 3.

By allowing the network to facilitate the use of multi-level of receptive field through the gradually increased global contexts of side inputs (Zhao et al., 2015; Liu et al., 2017), our proposed network can deal with unpredictable shapes and large size variations of AF pockets effectively.

2.2.3. Side-output layer

The side-outputs are produced from early layers in the decoding path as shown in Fig. 3. This approach is also inspired from M-net (Mehta and Sivaswamy, 2017).

In AF-net, the side outputs $\{\mathfrak{s}_{out}^{(i)}\}_{i=1}^{4}$ are produced from the last feature maps ($\boldsymbol{h}^{(17)}$, $\boldsymbol{h}^{(20)}$, $\boldsymbol{h}^{(23)}$, and $\boldsymbol{h}^{(27)}$) just before taking average un-pooling in the decoding path. The side outputs are computed as follows:

$$\begin{aligned}
\mathfrak{s}_{out}^{(1)} &= \sigma(\mathit{Ave\text{-}unpool}(\boldsymbol{h}^{(17)}, 8) \circledast \boldsymbol{W}^{\mathfrak{s}_{out}^{(1)}}),\\
\mathfrak{s}_{out}^{(2)} &= \sigma(\mathit{Ave\text{-}unpool}(\boldsymbol{h}^{(20)}, 4) \circledast \boldsymbol{W}^{\mathfrak{s}_{out}^{(2)}}),\\
\mathfrak{s}_{out}^{(3)} &= \sigma(\mathit{Ave\text{-}unpool}(\boldsymbol{h}^{(23)}, 2) \circledast \boldsymbol{W}^{\mathfrak{s}_{out}^{(3)}}),\\
\mathfrak{s}_{out}^{(4)} &= \sigma(\boldsymbol{h}^{(27)} \circledast \boldsymbol{W}^{\mathfrak{s}_{out}^{(4)}}).
\end{aligned} \tag{6}$$

where σ is sigmoid function, $\mathit{Ave\text{-}unpool}(\boldsymbol{x}, s)$ is the bilinear upsampling of a factor s to $\boldsymbol{x}$, and $\boldsymbol{W}^{\mathfrak{s}_{out}^{(i)}}$ is a set of 1×1 filters. By averaging all side-output maps, the final segmentation output is generated.

To train the network with the side-output maps, the loss function in (1) is defined by

$$\mathbb{L}_{seg}(f_{seg}(\boldsymbol{x}), \mathbf{T}) = \sum_{k=1}^{4} \mathbb{L}_{cross}(\mathfrak{s}_{out}^{(k)}(\boldsymbol{x}), \mathbf{T}), \tag{7}$$

where $\mathbb{L}_{cross}$ is the average cross-entropy loss, given by

$$\begin{aligned}
&\mathbb{L}_{cross}(\mathfrak{s}_{out}^{(k)}(\boldsymbol{x}), \mathbf{T}) =\\
&\quad -\frac{1}{|\Omega|} \sum_{(a_1,a_2)\in\Omega} \sum_{i=1}^{2} \mathbf{T}^{(i)}(a_1, a_2) \log((\mathfrak{s}_{out}^{(k)})^{(i)}(\boldsymbol{x})(a_1, a_2)).
\end{aligned} \tag{8}$$

Here, Ω is a pixel grid of input images $\boldsymbol{x}$ and $(\mathfrak{s}_{out}^{(k)})^{(i)}$ is an ith class output in $\mathfrak{s}_{out}^{(k)}$.

Since the side-output layers somewhat constrain the neural network and help train early layers by backpropagating the side-output loss to the early layers in the decoding path (Lee et al., 2015; Wang et al., 2015; Mehta and Sivaswamy, 2017), we expect the side output layers to prevent the propagation of redundant information for AF segmentation via the skip connections.

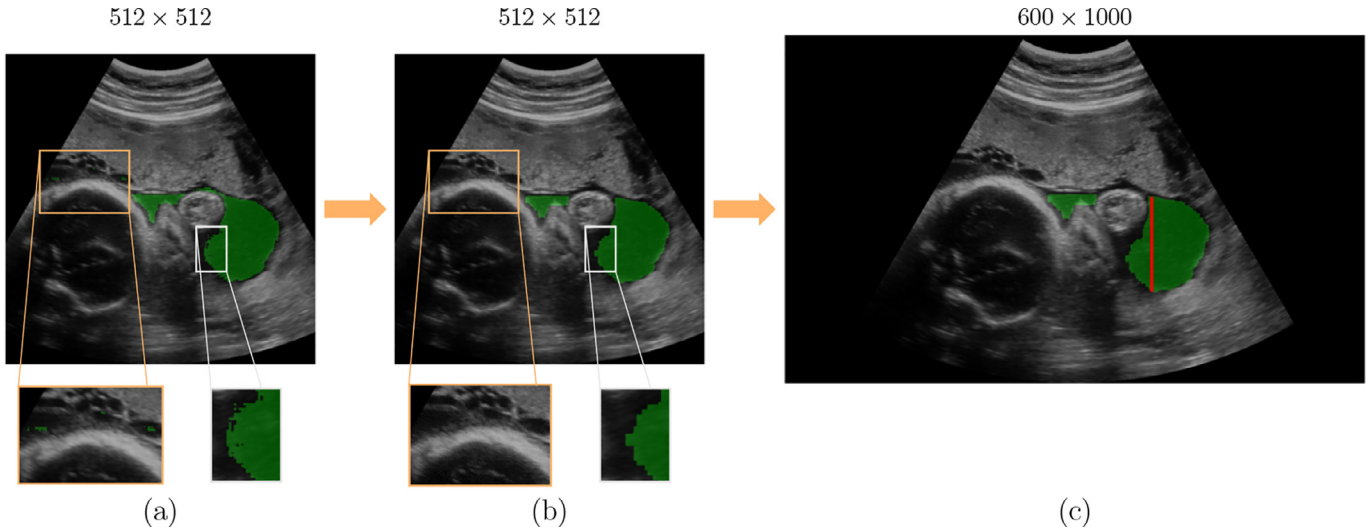


Fig. 8. Workflow of AFI measurement. For a given US image, we first obtain an estimated AF pocket region (green region in (a)) through a trained deep learning network. The segmentation result might include very small false positive regions (orange box in (a)) or several holes (white box in (a)). To fine-tune the segmentation result, we apply opening, followed by closing. These procedure can remove the morphological noises in the segmentation result, as shown in (b). Lastly, the result is rescaled into the original 600 × 1000 US image, as seen in (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2.4. Combination of three complementary techniques produce synergistic effect each others

AF-net is the optimal combination of three complementary techniques with U-net to deal with various confusing factors including reverberation, AF mimicking region, floating matter, and incomplete or missing boundary. Even experts can not identify the boundary of AF pocket by just looking at local region. Clinician uses considerable large global context and anatomical knowledge to recognize proper AF pocket. Our AF-net reflects the clinician's approach to AF pocket segmentation. To be precise, atrous convolution combines with side-input layers in the encoding path and side output layers in the decoding path to produce a synergistic effect with each other. This is because considerably enlarged receptive field of side-input layers and side-output layers via atrous convolution could effectively recognize proper AF pocket, in spite of confusing factors and unspecified size or shape of AF pocket.

2.3. Measurement of AFI

For the measurement of AFI, we apply the opening and closing method (Haralick et al., 1987) to the segmentation result, obtained from AF-net. A 10×10 square structuring element is used for the opening and closing of the binary image $T^{(1)}_{output}(\boldsymbol{x})$, which is the AF pocket segmentation output of $f_{seg}(\boldsymbol{x})$. Opening removes small false positive results from $A = \{T^{(1)}_{output}(\boldsymbol{x}) = 1\}$ and closing fills small holes in the opening of A, as described in Fig. 8. From the result of opening and closing, the deepest vertical length is obtained. Finally, the US image is rescaled to the original size and the pixel-wise length is converted to millimeter (mm).

3. Experiments and results

3.1. Data acquisition and experimental setting

For training and evaluation, 435 US images were provided by the Department of Obstetrics and Gynecology, Yonsei University College of Medicine, Seoul, Korea (IRB no.: 1-2019-0052). A total of 435 US images were obtained by expert sonographers with W10 and WS80A US machines (SAMSUNG Medison, Seoul, Korea) using a 2–6-MHz transabdominal transducer CA1-7A. The 435 US images were obtained from 255 patients. Pregnant women between $20+0$ and $36+6$ weeks of gestation was included. They were over 19 years old and Korean. Pregnant women with abnormally decreased amniotic fluid such as premature preterm rupture of membrane, and anhydramnios/oligohydramnios attributed in urinary abnormalities were excluded. We considered various conditions of each pregnant woman which could affect ultrasound image qualities such as location and size of AF pocket, subcutaneous fat distance, fat distribution, etc. The ground truth of AF pocket and AFI line were marked by two sonographers.

Table 1
Summary of datasets.

	Number of subjects	Number of images
Training set	155	310
Evaluation set	100	125
Total	255	435

For efficient training and computation, the original 1000×600 px US images were cropped into 600×600, removing unnecessary side parts, and then resized to 512×512 using bilinear interpolation. Three hundred and ten (310) US images from 155 patients were used for training neural networks and five-fold cross-validation was adopted in the training process, as depicted in Fig. 9. The remaining 125 US images from 100 patients were used for performance evaluation. Table 1 presents a summary of the datasets. Additionally, we augmented the training data with random contrast and brightness settings, and horizontal flipping to reduce the effects of overfitting and to make the network more robust and reliable.

Training processes were implemented using TensorFlow r1.13 (Girija, 2016) on a CPU (Intel CoreTM i7-6850 K, 3.60 GHz), 128 GB DDR4 RAM, and a four GPU system (NVIDIA GTX-1080ti, 11 GB). All weights were initialized by a zero-centered normal distribution with 0.01 standard deviation. The cost function was minimized using the AdamOptimizer (Kingma and Ba, 2014) with an initial learning rate of 0.0001, batch normalization (Ioffe and Szegedy, 2015), and mini-batch size of 4 at each epoch. To achieve better performance, we also used the "poly" learning rate policy (Liu et al., 2015; Chen et al., 2017a) where the learning rate at each epoch is chosen by a value multiplied by $(1 - \frac{epoch}{max\_epoch})^{power}$ to the initial learning rate with $power = 0.9$.

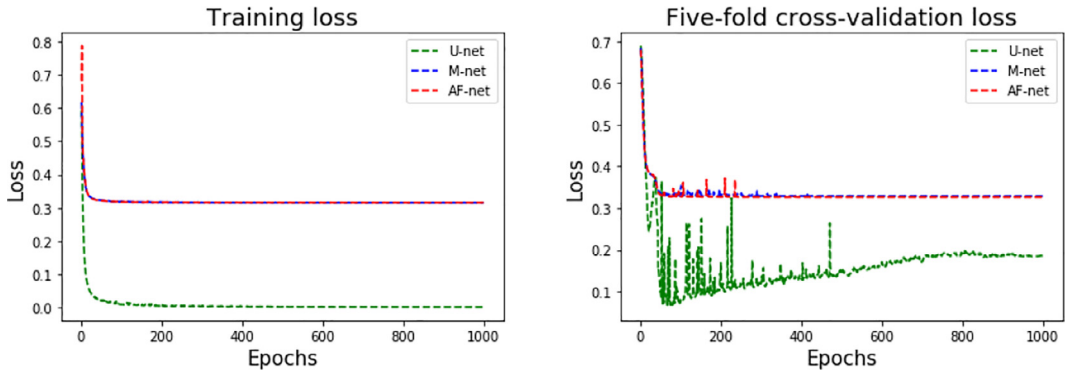


Fig. 9. Traces of the training loss (left) and five-fold cross-validation loss (right).

Table 2
Quantitative evaluation of 125 test segmentation results with four metrics; dice, precision, sensitivity, and specificity. The *p*-values are calculated by comparing AF-net with U-net and M-net ("n.s." stands for "not statistically significant").

	Dice		Precision		Sensitivity		Specificity	
	Mean ± std	*p*-value	Mean ± std	*p*-value	Mean ± std	*p*-value	Mean ± std	*p*-value
U-net	0.822 ± 0.157	1.8e−08	0.914 ± 0.092	n.s	0.782 ± 0.196	8.5e−13	0.995 ± 0.005	n.s
M-net	0.820 ± 0.168	1.0e−09	0.902 ± 0.098	n.s	0.791 ± 0.202	9.3e−11	0.994 ± 0.006	n.s
AF-net	0.877 ± 0.086	–	0.898 ± 0.111	–	0.873 ± 0.098	–	0.994 ± 0.008	–

3.2. Evaluate segmentation performance on AF-net with other models.

To assess the quality of AF pocket segmentation result, the following evaluation metrics were chosen; dice similarity, precision, sensitivity, and specificity (Udupa et al., 2006). Dice metric (d_D), which represents the degree of similarity between a segmented output and a ground-truth, is defined as

$$d_D := \frac{2|O_{\mathrm{GT}} \cap O_{\mathrm{AF}}|}{|O_{\mathrm{GT}}| + |O_{\mathrm{AF}}|}, \tag{9}$$

where O_{AF} is the AF pocket region obtained from a segmentation model and O_{GT} is the ground-truth annotated by an sonographer. The other metrics, namely precision(d_P), sensitivity(d_{Sen}), and specificity(d_{Spe}), which are defined by

$$d_P = \frac{|O_{\mathrm{GT}} \cap O_{\mathrm{AF}}|}{|O_{\mathrm{AF}}|}, d_{Sen} = \frac{|O_{\mathrm{GT}} \cap O_{\mathrm{AF}}|}{|O_{\mathrm{GT}}|} d_{Spe} = \frac{|(O_{\mathrm{GT}} \cup O_{\mathrm{AF}})^c|}{|(O_{\mathrm{GT}})^c|} \tag{10}$$

are obtained from the confusion matrix. In order to demonstrate the significant difference between AF-net and other models, a non-parametric Wilcoxon signed-rank test is applied on the evaluation results of AF pocket segmentation.

For the performance evaluation, the proposed AF-net is compared to the standard U-net and M-net. As indicated in Table 2, AF-net remarkably improved sensitivity, and dice similarity in terms of mean and standard deviation while maintaining precision and specificity. In terms of the *p*-value, the sensitivity and dice similarity of the models exhibited statistically significant differences, but precision and specificity did not. Qualitative evaluation of segmentation results in Fig. 10 shows that U-net and M-net often misrecognize an AF pocket due to confusing factors. On the other hand, our AF-net exhibits promising segmentation performance. Specifically, U-net and M-net fail to identify the proper boundaries of the AF pockets for cases (a), (c), and (d), while AF-net provides the promising segmentation result. For case (b) U-net and M-net misrecognize the AF mimicking region as an AF pocket, while AF-net successfully recognizes it. These segmented incorrect and isolated regions, which can lead to undesirable AFI measurement. On the other hand, AF-net provides accurate segmentation results without misrecognizing incorrect and isolated regions as AF pocket.

3.3. Effect of three complementary techniques on segmentation performance

To investigate the effects of three complementary techniques (atrous convolution, side-input layer, and side-output layer), we compare segmentation performance with U-net by adding several combinations of the complementary techniques (see Table 3 and Fig. 11). According to our experimental results, without atrous convolution, adding only either side-input layer or side-output layer or both does not improve the performance of AF pocket segmentation. In contrast, adding atrous convolution improves the segmentation performance. This is because a considerably enlarged receptive field size via atrous convolution can effectively recognize a proper AF pocket, despite the confusing factors and unspecified size or shape of the AF pocket. Atrous convolution combined with side-input layers in the encoding path and side-output layers in the decoding path to produce a synergistic effect with each other.

3.4. Evaluate AFI measurement performance on AF-net with other models.

In this subsection, we summarize evaluation results for final AFI measurement by comparing the accuracy with respect to ground truth. Let f_{seg} and f_{AFI} be a trained neural network for AF segmentation and a function of measurement of AFI from a segmentation result, respectively. To assess the quality of AFI measurement, the following evaluation metrics were used: Absolute error (d_{Ab}) and relative error (d_{Re}) are defined by

$$d_{Ab} = |f_{AFI}(f_{seg}(\boldsymbol{x})) - l|, \;\; d_{Re} = \frac{d_{Ab}}{l} \tag{11}$$

where $\boldsymbol{x}$ and l are an input US image and the corresponding AFI ground truth, respectively.

In Fig. 12, we provide quantitative evaluations of AFI measurement with three different AFI labels; a computed line and two annotated lines. As depicted in Fig. 13, the computed line is calculated by exactly the longest vertical line in the ground truth of AF

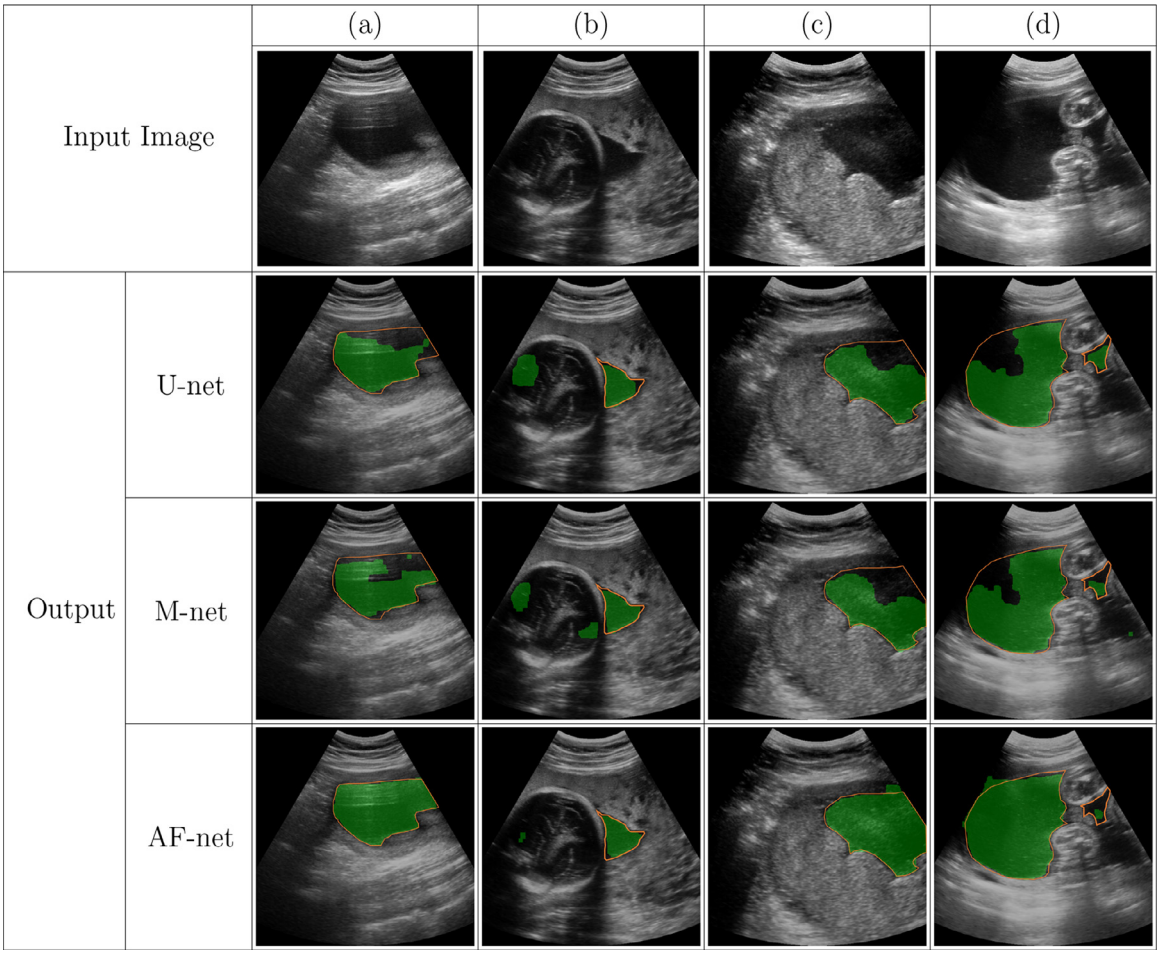


Fig. 10. Qualitative segmentation result comparison of U-net, M-net, and AF-net for four-cases with at least one dominate confusing factor((a) reverberation artifact, (b) AF mimicking region, (c) floating matters and (d) incomplete or missing boundary). As shown in (a), (c), and (d), the proposed AF-net provides suitable segmentation for the AF pocket despite the presence of confusing factors inside the AF pocket region.

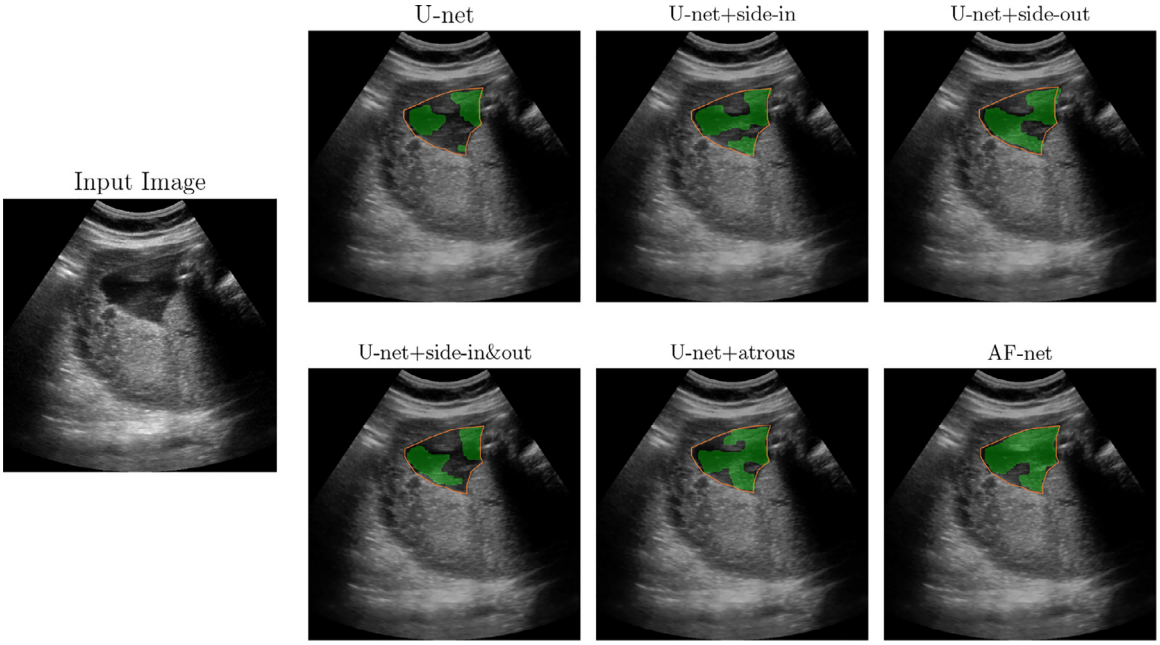
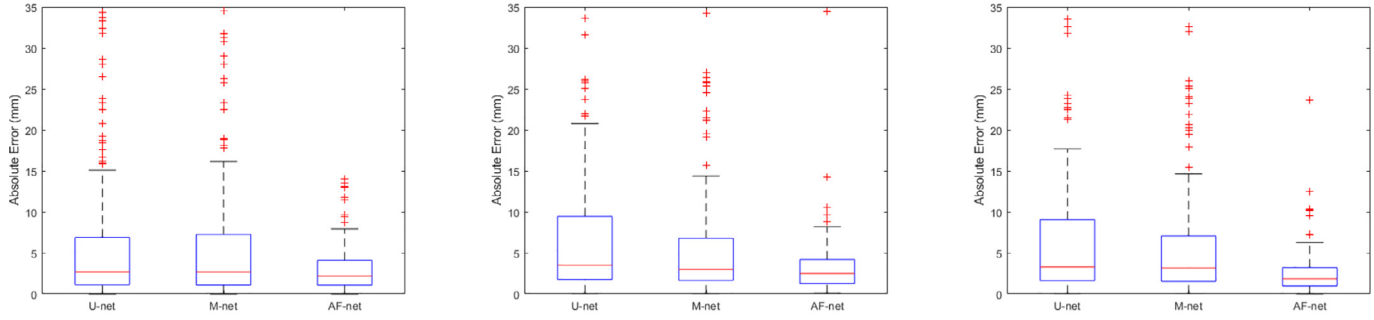


Fig. 11. Qualitative segmentation result comparison of complementary techniques for reverberation artifacts case.



(a) Computed line (b) Expert 1 (c) Expert 2

	Computed line				Expert1				Expert2			
	Absolute error(mm)		Relative error		Absolute error(mm)		Relative error		Absolute error(mm)		Relative error	
	Mean± std	p-value	Mean± std	p-value	Mean± std	p-value	Mean± std	p-value	Mean± std	p-value	Mean± std	p-value
U-net	7.550 ± 10.974	1.7e−5	0.038 ± 0.047	3.4e−5	7.40 ± 8.846	5.5e−8	0.041 ± 0.045	8.3e−8	7.101 ± 8.898	1.4e−8	0.040 ± 0.044	2.7e−8
M-net	6.592 ± 9.212	3.1e−5	0.036 ± 0.048	1.9e−4	6.178 ± 7.818	4.3e−5	0.038 ± 0.047	7.2e−5	6.252 ± 7.452	8.3e−8	0.039 ± 0.047	2.9e−7
AF-net	3.131 ± 3.032	–	0.019 ± 0.019	–	3.171 ± 3.699	–	0.021 ± 0.028	–	2.666 ± 2.986	–	0.018 ± 0.023	–

(d)

Fig. 12. Evaluation of 125 test results of AFI line estimation using three labels; computed lines by AFI definition and annotated lines by two experts. The figure includes box plots of absolute error in (a), (b) and (c) and a comparison table of absolute error and relative error in (d). The *p*-values are calculated by comparing AF-net with U-net and M-net ("n.s." stands for "not statistically significant").

Table 3
Quantitative evaluation of 125 test segmentation results for compare the effect of complementary techniques. The *p*-values are calculated by comparing AF-net with other models ("n.s." stands for "not statistically significant").

	Dice		Precision		Sensitivity		Specificity	
	Mean ± std	*p*-value	Mean ± std	*p*-value	Mean ± std	*p*-value	Mean ± std	*p*-value
U-net	0.822 ± 0.157	1.8e−08	0.914 ± 0.092	n.s	0.782 ± 0.196	8.5e−13	0.995 ± 0.005	n.s
U-net + side-in	0.813 ± 0.165	5.7e−10	0.906 ± 0.097	n.s	0.779 ± 0.205	5.2e−13	0.995 ± 0.006	n.s
U-net + side-out	0.818 ± 0.164	7.3e−10	0.912 ± 0.093	n.s	0.777 ± 0.197	7.3e−14	0.995 ± 0.005	n.s
U-net + side-in&out	0.820 ± 0.168	1.0e−09	0.902 ± 0.098	n.s	0.791 ± 0.202	9.3e−11	0.994 ± 0.006	n.s
U-net + atrous	0.862 ± 0.112	0.006	0.892 ± 0.114	n.s	0.856 ± 0.134	0.006	0.993 ± 0.009	n.s
AF-net (U-net + atrous + side-in&out)	0.877 ± 0.086	–	0.898 ± 0.111	–	0.873 ± 0.098	–	0.994 ± 0.008	–

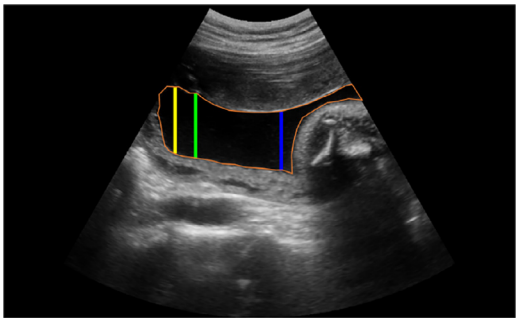


Fig. 13. Three AFI labels; computed lines (yellow line), Expert 1 annotations of AFI (blue line), and Expert 2 annotations of AFI (green line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

pocket, but the annotated lines are estimated by two experts via AFI measurement manual step. As seen in Fig. 12, AF-net achieves the smallest error range in measuring AFI with respect to all three different labels, compared to U-net and M-net.

As observed in Fig. 12(d), AF-net has the least mean and standard deviation of error for all three labels, and the *p*-value results demonstrate that all the error metrics exhibit statistically significant differences between models with respect to all the three labels. In addition, as shown in Fig. 12(a)–(c), AF-net provides more robust AFI measurement results, since there are only few outliers relative to U-net and M-net.

Fig. 14 illustrates qualitative evaluations for 5 different cases using the estimated AFI lines by the segmentation result of U-net, M-net, and AF-net. Compared to U-net and M-net, our proposed AF-net exhibits robust performance for all 5 cases.

U-net and M-net often do not provide a suitable AFI line when the ultrasound image is corrupted by various confusing factors or when there are large differences in the size of AF pockets. In Case 1, 3, 4, and 5, U-net and M-net produce unsuitable AFI lines owing to missing boundary, reverberation artifacts, and floating matters. In Case 2, U-net and M-net completely fail to obtain an AFI line because of the extremely small size of the AF pocket. In contrast, AF-net provides promising AFI lines.

The intra-observer and inter-observer measurements in Table 4 compare three AFI labels. The relatively large inter-observer variability between the two expert's annotations demonstrates that the detection of the deepest depth in an AF pocket somewhat relies on the sonographer's experience with a certain level of tolerance. As seen from the results in Fig. 12(d) and Table 4, the absolute and relative error of our proposed network with respect to three labels shows no significant difference compare to inter-observer variability between the labels. This means that our proposed method consistently provides an excellent candidate that is close to the AFI line.

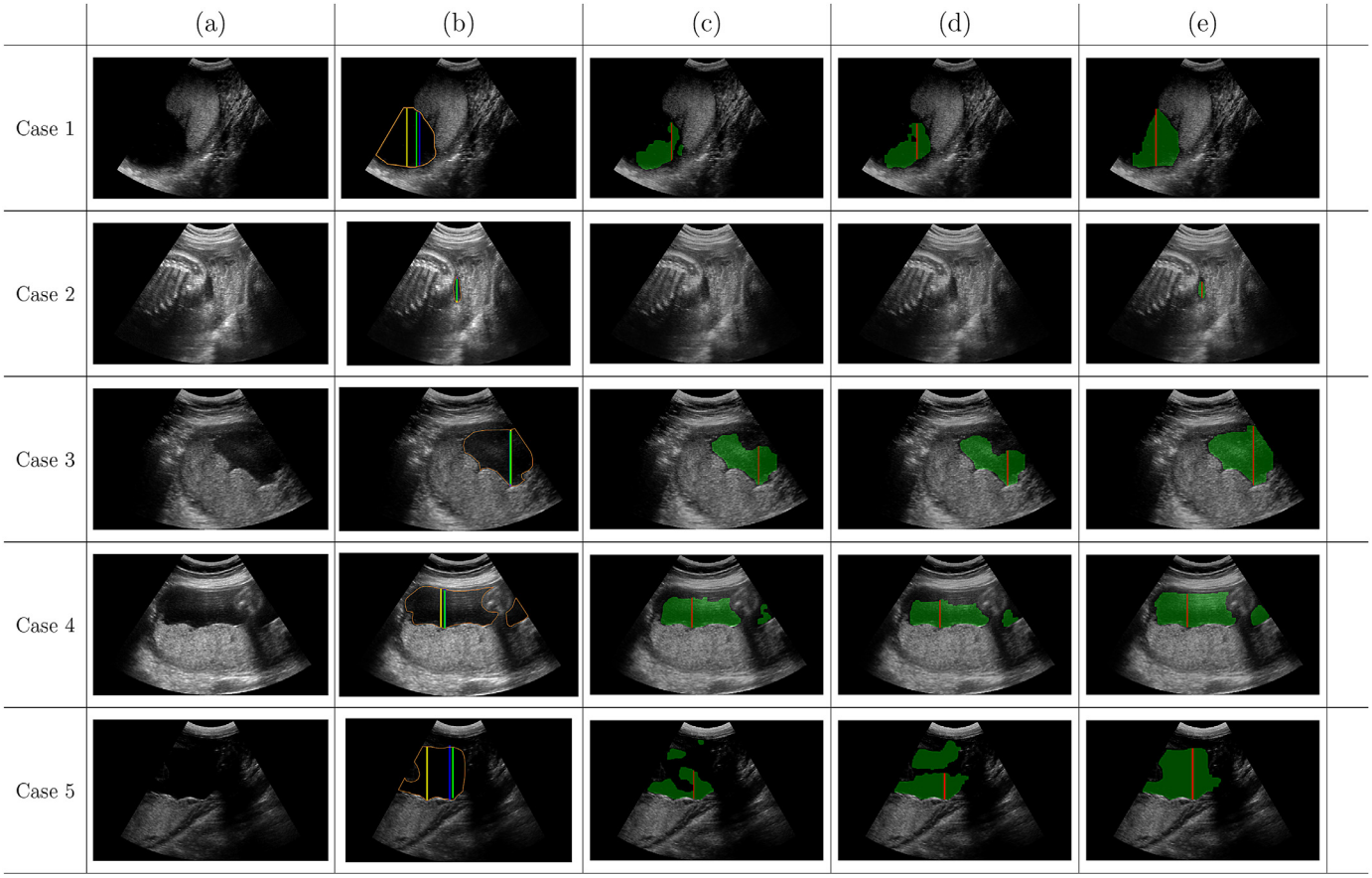


Fig. 14. Qualitative evaluation of AFI line estimation for five cases. (a) Original US images. (b) Ground truths of AF pocket boundary (orange line) with three labels; computed lines (yellow line), Expert 1 annotations of AFI (blue line), and Expert 2 annotations of AFI (green line). (c)–(e) Estimated AFI lines(red line) by AF pocket segmentation results(green region) using (c) U-net, (d) M-net, and (e) AF-net. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Intra-observer and inter-observer variability for three AFI line labels; the computed line by AFI definition and annotated lines by two experts.

Metric	Intra-observer		Inter-observer		
	Expert 1	Expert 2	Expert 1 & Expert 2	Expert 1 & computed line	Expert 2 & computed line
absolute error (mm)	2.348 ± 3.357	2.076 ± 1.930	3.772 ± 4.724	3.682 ± 5.274	3.629 ± 3.524
relative error	0.013 ± 0.018	0.014 ± 0.015	0.022 ± 0.027	0.023 ± 0.032	0.023 ± 0.024

4. Discussion and conclusion

This paper, for the first time, proposed a deep-learning-based method for automatic evaluation of AFI from ultrasound images. This task was extremely challenging, as it is quite difficult to segment accurately AF pockets due to amorphous features of AF pockets and ultrasound artifacts(i.e. reverberation, shadowing, speckle, signal dropout). Traditional segmentation techniques including shaped-based models may not be suitable to AF pocket segmentation. To be precise, most energy-based segmentation methods (using active contour or level-set) start from on initial contour and use the iterative contour evolution through energy minimization process. Therefore, these methods requires a good choice of initial contour and a suitable stopping criterion to terminate the iterative process at the target boundary. Unfortunately, in our AF pocket segmentation case, it is very difficult to find an energy functional satisfying the previously mentioned criteria, due to the confusing factors and amorphous feature of AF pocket as stated above. The proposed deep learning method is designed to address these difficulties by using three supplementary techniques in U-net.

The segmentation performance of the standard U-net appears to be fairly limited for our task. One main reason seems to be the significantly small size of the receptive field. Even the empirical receptive field size of U-net is much smaller than the theoretical one (Zhou et al., 2014). However, atrous convolution address this limitation by significantly enlarging the receptive size. The global context obtained from the large receptive size can be helpful for classifying an AF pocket region with the various ultrasound artifacts. In addition, the side input and output layers also seem to improve the AF pocket segmentation ability of U-net effectively through multi-level receptive field and prevention of useless information propagation in skip connections. Our experimental results validate advantages of the supplementary techniques in detecting an AF pocket even for US images that include various ultrasound artifacts, small dark regions resembling AF, or large variations of AF pocket size.

The major benefit of automated AFI measurement is to minimize not only the sonographers' workload but also the inter-variability or the risk of human error by providing a high quality AFI line candidate close to the ideal line in the sense of the defini-

tion of AFI. Because AFI line measurement is highly dependent on the sonographers experience, there have been considerable differences between the annotated lines by two sonographers even with significant clinical experience.

The proposed method has rooms for further improvement. It would be desirable to deal with uncertain factors such as direction and angles of the transducer and pressure with the transducer. With the patient supine, the linear or curvilinear transducer is placed along the maternal anterior abdominal wall and held perpendicular to the floor. A few potential pitfalls were mentioned in the textbook (Norton, 2016) during determining an estimated AFV. Maternal position should be supine and at least within 45-degree semi-erect position during measuring AFV (Tressler et al., 2006). And Fok et al. (2006) was demonstrated that fetal position could influence the calculation of AFI although AFV is not altered by fetal position. This is due to the direction and angles of the transducer. Excessive pressure with the transducer also affect estimation of AFV. Flack was reported that excessive pressure on maternal abdomen resulted in 21% decrease in calculated AFI measurement (Flack et al., 1994). Moreover, ultrasound beam is scattered in adipose tissue and resulted in artificial eches. Hence, obese women may appear to have reduced amniotic fluid. A lot of variation lies in estimation of AFV for these various conditions, which leads to time consuming. Our future research direction is to develop a consistent, accurate, and robust method to handle various difficult conditions mentioned above.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Hyun Cheol Cho: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Siyu Sun:** Software, Validation, Visualization, Formal analysis, Investigation. **Chang Min Hyun:** Writing - original draft, Writing - review & editing. **Ja-Young Kwon:** Supervision, Project administration, Funding acquisition, Data curation. **Bukweon Kim:** Writing - original draft, Resources. **Yejin Park:** Supervision, Project administration, Data curation, Writing - original draft, Writing - review & editing. **Jin Keun Seo:** Supervision, Project administration, Funding acquisition, Writing - original draft.

Acknowledgments

This work was supported by Samsung Medison and Samsung Science & Technology Foundation (No. SSTF-BA1402-01). C.H.C. and J.K.S. were supported in part by the National Research Foundation of Korea (NRF) Grant 2015R1A5A1009350 and 2017R1A2B20005661. We would like to express our deepest gratitude to two sonographers who make a ground truth data, Hye Mi Jeon and Hye Ri Kim. We also appreciate Haeeun Han for helping to draw Fig. 1

References

Beall, M., Van Den Wijngaard, J., Van Gemert, M., Ross, M., 2007. Amniotic fluid water dynamics. Placenta 28 (8-9), 816–823.

Chamberlain, P., Manning, F., Morrison, I., Harman, C., Lange, I., 1984. Ultrasound evaluation of amniotic fluid volume: I. The relationship of marginal and decreased amniotic fluid volumes to perinatal outcome. Am. J. Obstet. Gynecol. 150 (3), 245–249.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017b. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.

Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y., 2017. Automatic brain tumor detection and segmentation using U-net based fully convolutional networks. In: Annual Conference on Medical Image Understanding and Analysis. Springer, pp. 506–517.

Dubil, E.A., Magann, E.F., 2013. Amniotic fluid as a vital sign for fetal wellbeing. Aust. J. Ultrasound Med. 16 (2), 62–70.

Flack, N.J., Doré, C., Southwell, D., Kourtis, P., Sepulveda, W., Fisk, N.M., 1994. The influence of operator transducer pressure on ultrasonographic measurements of amniotic fluid volume. Am. J. Obstet. Gynecol. 171 (1), 218–222.

Fok, W., Chan, L., Lau, T., 2006. The influence of fetal position on amniotic fluid index and single deepest pocket. Ultrasound Obstet. Gynecol. 28 (2), 162–165.

Girija, S. S., 2016. Tensorflow: large-scale machine learning on heterogeneous distributed systems. Software available from tensorflow. org 39.

Haralick, R.M., Sternberg, S.R., Zhuang, X., 1987. Image analysis using mathematical morphology. IEEE Trans. Pattern Anal. Mach. Intell. (4) 532–550.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–19165.

Hu, Y., Guo, Y., Wang, Y., Yu, J., Li, J., Zhou, S., Chang, C., 2019. Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. Med. Phys. 46 (1), 215–228.

Hyun, C.M., Kim, H.P., Lee, S.M., Lee, S., Seo, J.K., 2018. Deep learning for undersampled MRI reconstruction. Phys. Med. Biol. 63 (13), 135007.

Hyun, C. M., Kim, K. C., Cho, H. C., Cho, J. K., Seo, J. K., 2019. Framelet pooling aided deep learning network: the method to process high dimensional medical data. arXiv:1907.10834.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167

Jang, J., Park, Y., Kim, B., Lee, S.M., Kwon, J.-Y., Seo, J.K., 2017. Automatic estimation of fetal abdominal circumference from ultrasound images. IEEE J. Biomed. Health Inform. 22 (5), 1512–1520.

Kehl, S., Schelkle, A., Thomas, A., Puhl, A., Meqdad, K., Tuschy, B., Berlit, S., Weiss, C., Bayer, C., Heimrich, J., et al., 2016. Single deepest vertical pocket or amniotic fluid index as evaluation test for predicting adverse pregnancy outcome (safe trial): a multicenter, open-label, randomized controlled trial. Ultrasound Obstet. Gynecol. 47 (6), 674–679.

Kim, B., Kim, K.C., Park, Y., Kwon, J.-Y., Jang, J., Seo, J.K., 2018. Machine-learning-based automatic identification of fetal abdominal circumference from ultrasound images. Physiol. Meas. 39 (10), 105007.

Kim, H.P., Lee, S.M., Kwon, J.-Y., Park, Y., Kim, K.C., Seo, J.K., 2019. Automatic evaluation of fetal head biometry from ultrasound images using machine learning. Physiol. Meas. 40 (6), 065009.

Kingma, D. P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv:1412.6980

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z., 2015. Deeply-supervised nets. In: Artificial Intelligence and Statistics, pp. 562–570.

Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T., 2019. Deep learning in medical ultrasound analysis: a review. Engineering 5 (2), 261–275.

Liu, W., Rabinovich, A., Berg, A. C., 2015. Parsenet: looking wider to see better. arXiv:1506.04579

Liu, Y., Cheng, M.-M., Hu, X., Wang, K., Bai, X., 2017. Richer convolutional features for edge detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3000–3009.

Luntsi, G., Burabe, F.A., Ogenyi, P.A., Zira, J.D., Chigozie, N.I., Nkubli, F.B., Dauda, M., 2019. Sonographic estimation of amniotic fluid volume using the amniotic fluid index and the single deepest pocket in a resource-limited setting. J. Med. Ultrasound 27 (2), 63.

Magann, E.F., Chauhan, S.P., Doherty, D.A., Lutgendorf, M.A., Magann, M.I., Morrison, J.C., 2007. A review of idiopathic hydramnios and pregnancy outcomes. Obstet. Gynecol. Surv. 62 (12), 795–802.

Magann, E.F., Sandlin, A.T., Ounpraseuth, S.T., 2011. Amniotic fluid and the clinical relevance of the sonographically estimated amniotic fluid volume: oligohydramnios. J. Ultrasound Med. 30 (11), 1573–1585.

Manning, F.A., Platt, L.D., Sipos, L., 1980. Antepartum fetal evaluation: development of a fetal biophysical profile. Am. J. Obstet. Gynecol. 136 (6), 787–795.

Mehta, R., Sivaswamy, J., 2017. M-net: a convolutional neural network for deep brain structure segmentation. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, pp. 437–440.

Moore, T.R., Cayle, J.E., 1990. The amniotic fluid index in normal human pregnancy. Am. J. Obstet. Gynecol. 162 (5), 1168–1173.

Norton, M., 2016. Callen's Ultrasonography in Obstetrics and Gynecology, sixth ed. Elsevier.

of ObstetriciansA.C., Gynecologists, 2014. Practice bulletin no. 145: antepartum fetal surveillance. Obstet. Gynecol. 124 (145), 182–192.

of ObstetriciansA.C., Gynecologists, the American Institute of Ultrasound in Medicine, 2016. Practice bulletin no. 175 summary: ultrasound in pregnancy. Obstet. Gynecol. 128 (175), e241–e256.

Phelan, J., Ahn, M., Smith, C.V., Rutherford, S., Anderson, E., 1987. Amniotic fluid index measurements during pregnancy.. J. Reprod. Med. 32 (8), 601–604.

Phelan, J., Smith, C.V., Broussard, P., Small, M., 1987. Amniotic fluid volume assessment with the four-quadrant technique at 36–42 weeks' gestation.. J. Reprod. Med. 32 (7), 540–542.

Pri-Paz, S., Khalek, N., Fuchs, K., Simpson, L., 2012. Maximal amniotic fluid index as a prognostic factor in pregnancies complicated by polyhydramnios. Ultrasound Obstet. Gynecol. 39 (6), 648–653.
Rabie, N., Magann, E., Steelman, S., Ounpraseuth, S., 2017. Oligohydramnios in complicated and uncomplicated pregnancy: a systematic review and meta-analysis. Ultrasound Obstet. Gynecol. 49 (4), 442–449.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.
Rutherford, S.E., Phelan, J.P., Smith, C.V., Jacobs, N., 1987. The four-quadrant assessment of amniotic fluid volume: an adjunct to antepartum fetal heart rate testing.. Obstet. Gynecol. 70 (3 Pt 1), 353–356.
Tressler, T., Bernazzoli, M., Hole, J., Martinez, F., 2006. The effects of maternal position on the amniotic fluid index. J. Ultrasound Med. 25 (4), 445–447.
Udupa, J.K., LeBlanc, V.R., Zhuge, Y., Imielinska, C., Schmidt, H., Currie, L.M., Hirsch, B.E., Woodburn, J., 2006. A framework for evaluating image segmentation algorithms. Comput. Med. Imaging Graph. 30 (2), 75–87.
Wang, L., Lee, C.-Y., Tu, Z., Lazebnik, S., 2015. Training deeper convolutional networks with deep supervision. arXiv:1505.02496
Wu, C.-S., Chen, C.-M., Chou, H.-C., 2017. Pulmonary hypoplasia induced by oligohydramnios: findings from animal models and a population-based study. Pediatri. Neonatol. 58 (1), 3–7.
Wu, L., Cheng, J.-Z., Li, S., Lei, B., Wang, T., Ni, D., 2017. Fuiqa: fetal ultrasound image quality assessment with deep convolutional networks. IEEE Trans. Cybern. 47 (5), 1336–1349.
Yap, M.H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., Davison, A.K., Martí, R., 2017. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J. Biomed. Health Inform. 22 (4), 1218–1226.
Yefet, E., Daniel-Spiegel, E., 2016. Outcomes from polyhydramnios with normal ultrasound. Pediatrics 137 (2), e20151948.
Zhao, R., Ouyang, W., Li, H., Wang, X., 2015. Saliency detection by multi-context deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1265–1274.
Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2014. Object detectors emerge in deep scene CNNs. arXiv:1412.6856